Project N^o: **FP7-284731**

Project Acronym: **UaESMC**

Project Title: **Usable and Efficient Secure Multiparty Computation**

Instrument: **Specific Targeted Research Project**

Scheme: **Information & Communication Technologies**

Future and Emerging Technologies (FET-Open)

Deliverable D1.1
Capability Model

Due date of deliverable: 31st March 2012

Actual submission date: 31st March 2012

Start date of the project: **1st February 2012** Duration: **36 months**

Organisation name of lead contractor for this deliverable: **UT**

Specific Targeted Research Project supported by the 7th Framework Programme of the EC		
Dissemination level		
PU	Public	✓
PP	Restricted to other programme participants (including Commission Services)	
RE	Restricted to a group specified by the consortium (including Commission Services)	
CO	Confidential, only for members of the consortium (including Commission Services)	

Executive Summary:

Capability Model

This document summarizes deliverable D1.1 of project FP7-284731 (UaESMC), a Specific Targeted Research Project supported by the 7th Framework Programme of the EC within the FET-Open (Future and Emerging Technologies) scheme. Full information on this project, including the contents of this deliverable, is available online at `http://www.usable-security.eu`.

The report contains a guide to the possible working applications of secure multi-party computation (SMC) through introduction of various models. Each of the models is described through key components, a visual image and a short descriptive example of a potential application. The models will be used to convey the key ideas of SMC to parties not familiar to the concept. As such, this deliverable forms the basis of task 1.2 where possible problems and needs to solve with the help of SMC will be sourced from different communities.

The goal of this deliverable is to help conveying the idea of SMC. As such, it concentrates on a particular implementation and security model, and does not aim to give an overview of different set-ups (number and roles of parties, details of the security model) of SMC.

List of authors:

Pille Pruulmann-Vengerfeldt

Liina Kamm

Riivo Talviste

Peeter Laud

Dan Bogdanov

Contents

1 Guide to Understanding Secure Multi-party Computation: Introduction

This is a guide to the possible working applications of secure multi-party computation (SMC) through introduction of various models. Each of the models is described through key components, a visual image and a short descriptive example of a potential application.

2 SMC Model Components

Data miner is a party who hosts data and performs computational tasks in conjunction with others.

Data provider is a party who provides data over a secure channel.

Data owner is a party who owns the provided data. This party may be a statistics office gathering the data of individuals or it may be the data provider itself who also owns the data. For model purposes, data providers and owners are only separated if they are not the same party.

Secure multi-party computation is a cryptographic technique allowing the owners of data to make it available as inputs of a computation in a manner that in the end of that computation, each party learns only the output assigned to it (and everything deducible from its inputs and outputs), but nothing more. Computation is based on encrypted data and, hence, data are never decrypted in order to compute, making records inaccessible to the computing parties. In all cases that we handle in this model, SMC signifies distributed data storage and computing for gathering and analysing delicate data without finding out what the individual values are.

In the most general form, **the data** on which the computation is performed can be imagined to be represented as a big table, where the rows correspond to individual cases that we are handling (e.g. Person1, Person2, Person3 or Company1, Company2) and columns correspond to different data that are known about those cases (e.g. Age, Gender, Salary or Resource1, Resource2). Each data provider contributes values for some cells of the table. There are two basic ways in which the data may be partitioned among the different data providers: horizontally or vertically (hybrid cases are possible, too, but significantly less common). **Horizontal partitioning** means that each provider provides all information about some of the cases (i.e. rows). **Vertical partitioning** means that each provider provides certain information about all cases (i.e. columns).

It is possible to build in **automatic checks** of data, but it does not eliminate the possibility for a data provider to lie or cheat. The aim of the model is to prevent data miners from looking at individual values, but it assumes that all participants are honest, but curious and that cooperation through correct data outweighs the benefits of cheating. To some extent this can be analysed using game theory in each particular case. Also, the additional layer of anonymity and security provides less incentive to lie.

While SMC prevents looking at individual values, there is still a possibility to check that the data is in a certain range (e.g. a person's age is not over 200). However, data owners must agree upon these checks so that this feature is not used maliciously.

There are dimensions of trust and dimensions of financial viability in each case and these can be discussed during the interviews.

3 Cases

We discuss the following cases:

- Case A: Three competitors
- Case B1: Two competitors and a "neutral" computer
- Case B2: Several energy market suppliers and several customers in joint interest finding
- Case B3: Several energy market suppliers and one customer OR one supplier and several customers
- Case C: An organization interested in its members' info
- Case D1: A researcher hosting sensitive data in a cloud
- Case D2a: A researcher interested in data from state databases using third-party miners
- Case D2b: A researcher interested in data from state databases using built-in computing possibilities of state infrastructure
- Case D3: Statistical data collection organizations working together for better results
- Case D4: Statistics office collecting data with improved security
- Case E1: State database and interested non-state parties
- Case E2: State database and interested non-state parties (with the computation outsourced)

3.1 Case A: Three competitors

Who are data providers?	All three parties
Who are data miners?	All three parties
What are the challenges for the trust?	Two parties working together, however, they would need to disclose their own data as well.
How is the data collated?	Horizontally
Who can submit the query?	All three parties
How is responsibility for the query distributed?	All three parties have agreed for joint query – no single party can perform any query alone
Who get to know the results?	Only the involved parties
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel.
What is the motivation for computing and keeping the trust?	The benefit from computation is bigger than the benefit from breaking the trust. Jointly using the confidential data in the interest of all can be more beneficial than getting caught learning each other's secrets.
Illustration	Figure 1

Three small-size companies decide to join forces and make a joint bid for a larger project. However, they need to calculate resources, revenues and input, without disclosing too much information. SMC enables them to share enough information for a joint project, but keep enough private to continue competing in other projects.

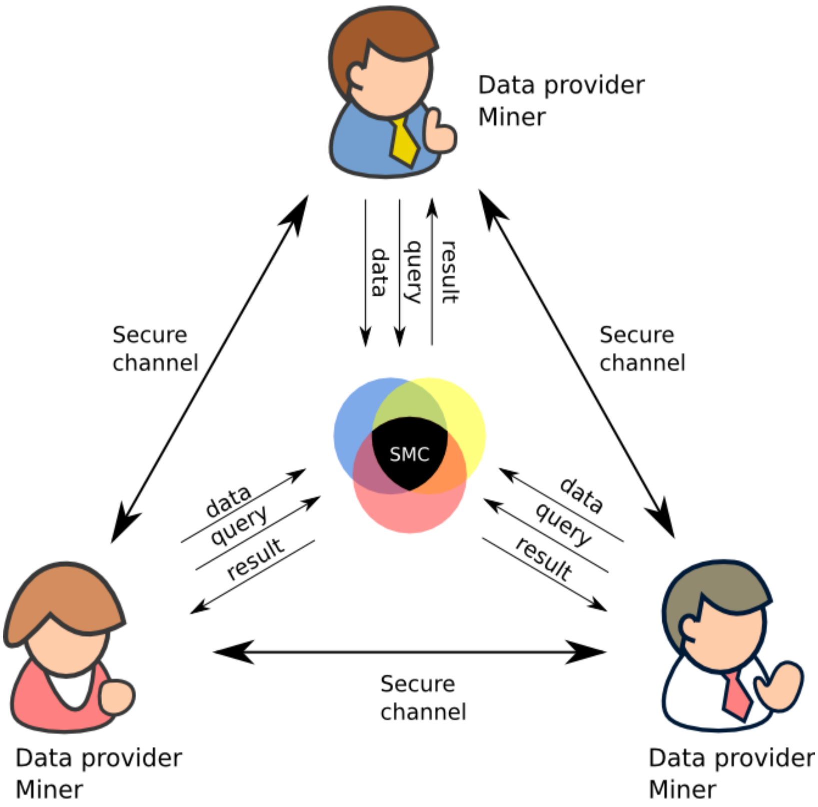

Figure 1. Case A: Three competitors

3.2 Case B1: Two competitors and a "neutral" computer

Who are data providers?	Two parties
Who are data miners?	The two data providers are miners themselves with assistance from the third neutral miner
What are the challenges for the trust?	The challenge is to trust the third party who might be working together with your competitor Data shuffling should help to keep each provider's data private.
How is the data collated?	Horizontally
Who can submit the query?	Data providers by joint agreement working together with the third party
How is responsibility for the query distributed?	Data providers by joint agreement
Who get to know the results?	Only the data providers
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel.
What is the motivation for computing and keeping the trust?	Two have put their data at stake, while the third, neutral party puts its trustworthiness and reputation at stake
Illustration	Figure 2

In principle, this model works in the case of two competitors wanting to figure out where is their optimal point of interest. For instance if a farmer has a supply curve and a producer has a demand curve, SMC helps them in figuring out their optimal price-production ratio. Additional information on auctions and supply and demand curve is explainable by game theory, where the incentives and motivations are thoroughly investigated.

The other "classical" example for this case is two rich people wanting to know who is richer, without disclosing their actual worth to the competitor. By entering their "value" to SMC model, they are able to compare their worth by involving a third party, but never disclosing their actual values.

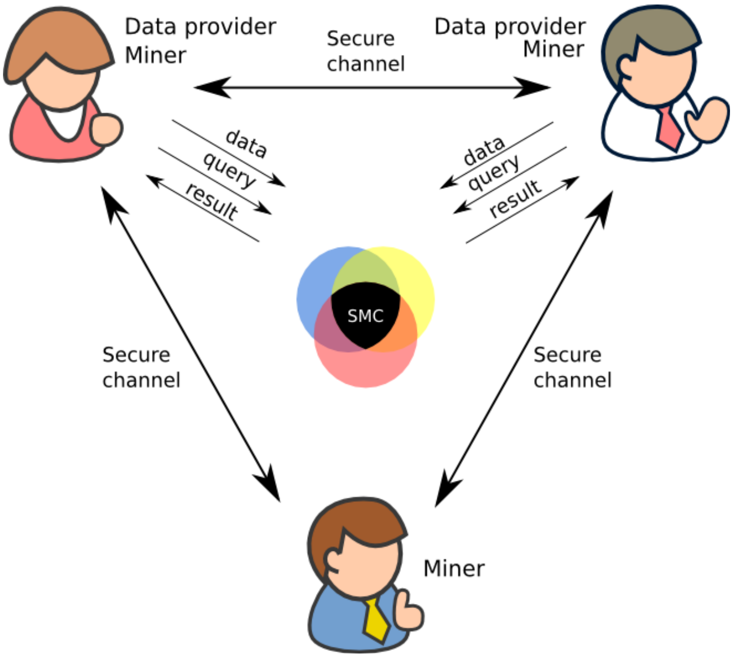

Figure 2. Case B1: Two competitors and a "neutral" computer

3.3 *Case B2: Several energy market suppliers and several customers in joint interest finding*

Who are data providers?	Numerous parties both on supplier and customer side
Who are data miners?	Both supplier and customer side data providers select amongst themselves representatives as miners and find the third trusted miner from outside. Or if they do not have mining capabilities, they can outsource mining to three different competitors.
What are the challenges for the trust?	The challenge includes cooperation of miners on supply and demand side or cooperation with the third party miners
How is the data collated?	Horizontally
Who can submit the query?	Data providers both on supplier and customer side by joint agreement or through elected representatives
How is responsibility for the query distributed?	Customer or suppliers as data providers by joint agreement
Who get to know the results?	Only the elected representatives who submitted the query
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel
What is the motivation for computing and keeping the trust?	Gains for working together outweigh the incentives to cheat
Illustration	Figure 3

Several distributed small power generators hold their power capabilities/costs (e.g. maximum power per hour, cost per watt per hour) and several consumers hold their power requirements (e.g. schedule constraints of devices and their power requirements). The consumers want to calculate their power allocations in order to minimize the total cost. However, the consumers do not want to disclose their allocation/requirements and generators do not want to disclose their costs. Game theory can be exploited to provide further constraints and assure that customers do not lie (e.g. reducing total costs implies reducing customer price).

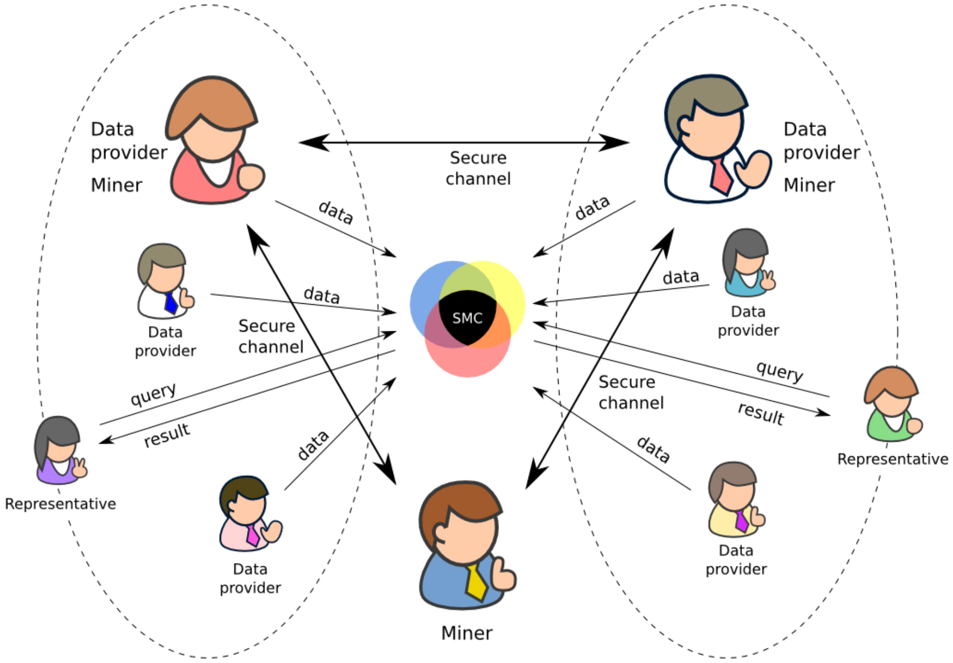

Figure 3. Case B2: Several energy market suppliers and several customers in joint interest finding

3.4 Case B3: Several energy market suppliers and one customer OR one supplier and several customers

Who are data providers?	Numerous parties both on supplier side and one customer (or vice versa)
Who are data miners?	All supplier side data providers select one trusted miner, customer side provides one data miner and the third miner is a neutral party. Or if they do not have mining capabilities, they can outsource mining to three different competitors.
What are the challenges for the trust?	The challenge includes cooperation of miners on the supply and demand side or cooperation with the third party miners.
How is the data collated?	Horizontally
Who can submit the query?	Data providers both on supplier and customer side by joint agreement or through elected representatives (e.g. organization's board)
How is responsibility for the query distributed?	Customer or suppliers as data providers by joint agreement and later by elected representatives
Who get to know the results?	Only the elected representatives who submitted the query
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel
What is the motivation for computing and keeping the trust?	Gains for working together outweigh the incentives to cheat
Illustration	Figure 4

Several distributed small power generators their power capabilities/costs (e.g. maximum power per hour, cost per watt per hour) and a consumer holds its power requirements (e.g. schedule constraints of devices and their power requirements). The consumer wants to calculate its power allocations in order to minimize the total cost. The consumer is not interested in disclosing their allocation/requirements and generators do not want to disclose their costs.

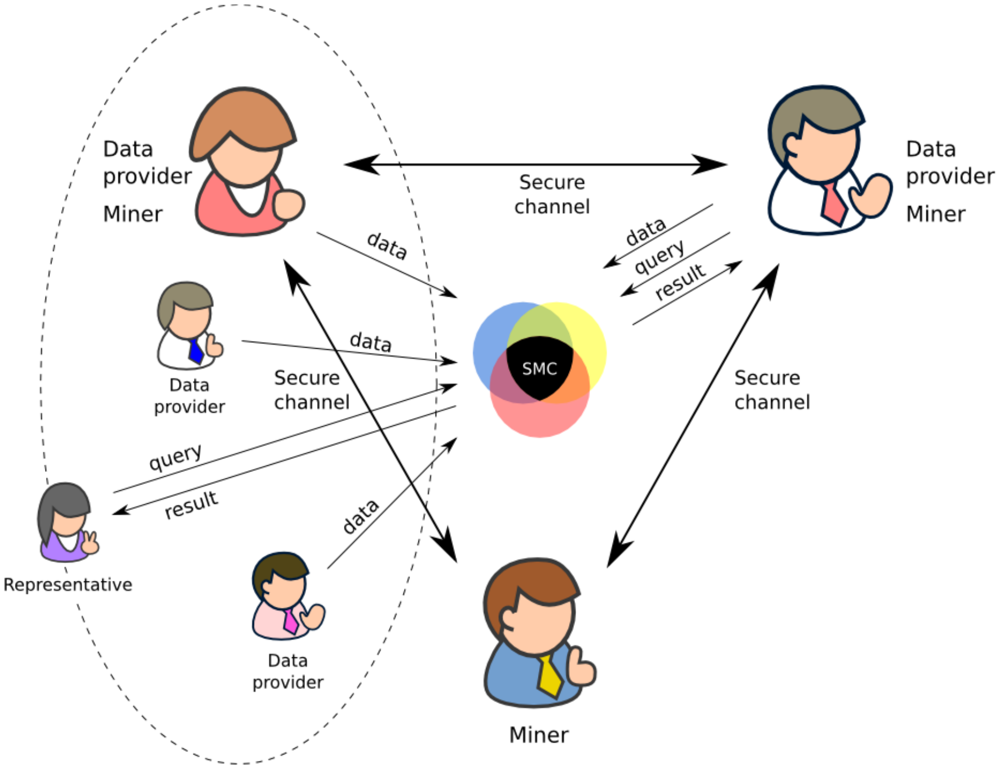

Figure 4. Case B3: Several energy market suppliers and one customer

3.5 Case C: An organization interested in its members' info

Who are data providers?	Numerous parties
Who are data miners?	Data providers select amongst themselves three trusted miners. Or if they do not have mining capabilities, they can outsource mining to three different competitors.
What are the challenges for the trust?	Two parties working together, however, they would need to disclose their own data as well. Data shuffling helps avoid two parties working together to figure out everyone else's info.
How is the data collated?	Horizontally
Who can submit the query?	Data providers by joint agreement or through elected representatives (e.g. organization's board)
How is responsibility for the query distributed?	All data providers by joint agreement
Who get to know the results?	Only the data providers or only the elected representatives who submitted the query.
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel.
What is the motivation for computing and keeping the trust?	All miners put their trustworthiness and reputation at stake among their peers
Illustration	Figure 5. The circle depicts numerous data providers belonging to the same organization

Here an organization with a number of members is interested in its members' fiscal data in a faster and more accurate manner than statistics office is able to provide. This is an actually realised case with Estonian Association of Information Technology and Telecommunications.

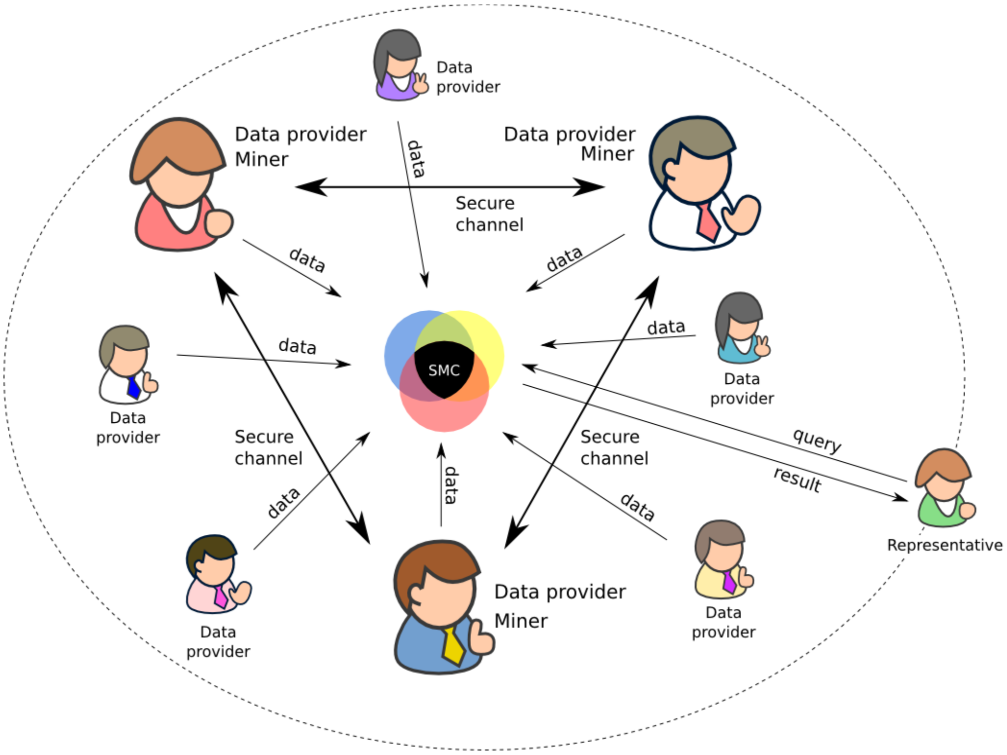

Figure 5. Case C: An organization interested in its members' info

3.6 Case D1: A researcher hosting sensitive data in a cloud

Who are data providers?	Although in principle there are numerous data providers (e.g. survey respondents or doctor's patients), their individual interest in their data is relatively low. Thus, for the case of cloud SMC, the data provider can also be considered to be one party – data owner – responsible for collecting the data and maintaining the collections
Who are data miners?	Data owner selects three SMC capable competing cloud service providers
What are the challenges for the trust?	Two parties working together. For data providers, the trust is mainly in the data owner
How is the data collated?	Horizontally
Who can submit the query?	Data owner
How is responsibility for the query distributed?	Data owner
Who get to know the results?	Only the data owner, responsible for data collection and organization
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel; Insecure data input channels
What is the motivation for computing and keeping the trust?	The idea of using SMC provides additional security for the data providers to increase their willingness to provide data. Also, in certain cases, it is financially sensible to keep your data securely in a set of clouds rather than build your own server capabilities. Data security officers need to approve of this solution beforehand
Illustration	Figure 6. The clouds depict numerous SMC capable cloud-computers

A researcher is interested in conducting survey on sensitive data – e.g. sexual encounters, health information or so. The trust-worthiness of the survey comes from security assurance that the data collected will be computed anonymously, also the data will be stored securely, thus it is very hard to get to the sensitive data.

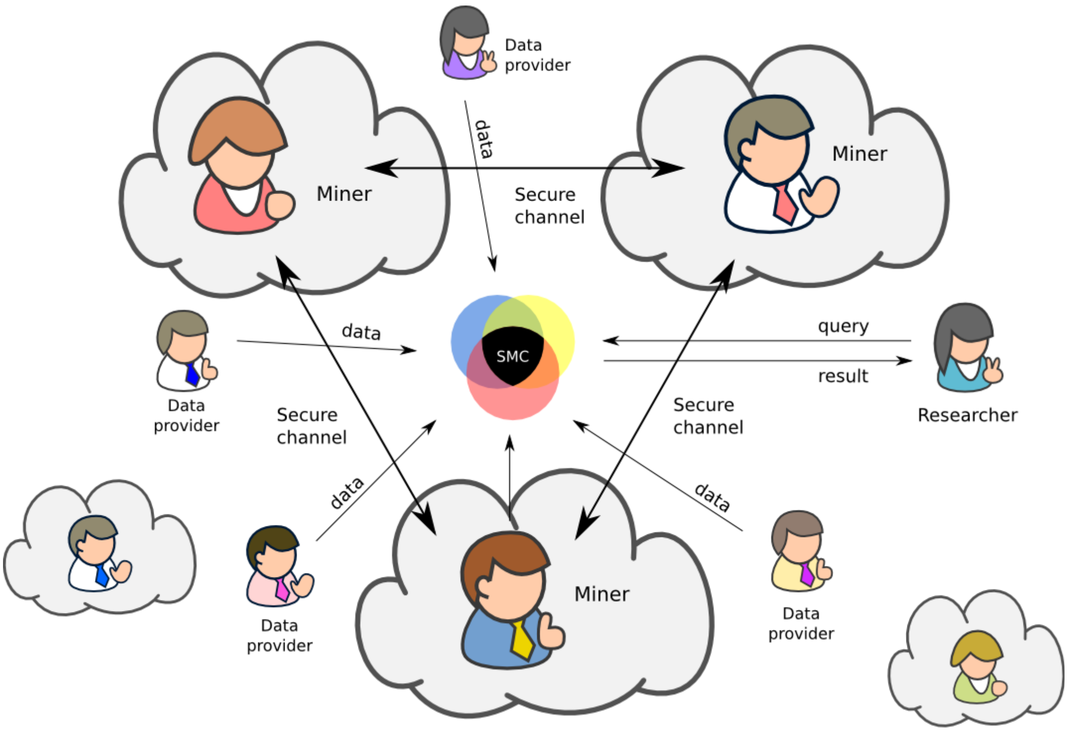

Figure 6. Case D1: A researcher hosting sensitive data in a cloud

3.7 Case D2a: A researcher interested in data from state databases using third-party miners

Who are data providers?	In this case, there are two layers of data providers. For the SMC, the data are provided from state owned databases. Each of these has also numerous data providers whose interest is in maximum protection of their information and proper data protection laws being reinforced.
Who are data miners?	State has three trusted SMC computing capable servers or trusted outside parties
What are the challenges for the trust?	Challenges for the trust are mostly in proper query control
How is the data collated?	Vertically
Who can submit the query?	Interested researcher from outside
How is responsibility for the query distributed?	Legislatively regulated responsibility for checking the adequacy of the query
Who get to know the results?	The interested researcher and checking authorities.
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel
What is the motivation for computing and keeping the trust?	SMC enables to put state databases into maximum use for research or policy making purposes, while maintaining security and privacy of each database and each data provider for those databases and avoiding merge of actual data. Data security officers need to approve of the security and legality of the solution.
Illustration	Figure 7. The registry information aggregation service is state owned infrastructure for secure data aggregation

State owns numerous databases, but the state is legally not allowed to merge the databases for any joint computation. We look at an interested researcher or policy maker, who is with some authorization, allowed to make a query that gives a general answer across multiple databases, but computations that enable case-based identification are not allowed. The computations are either conducted on state owned SMC capable servers or outsourced to trusted parties.

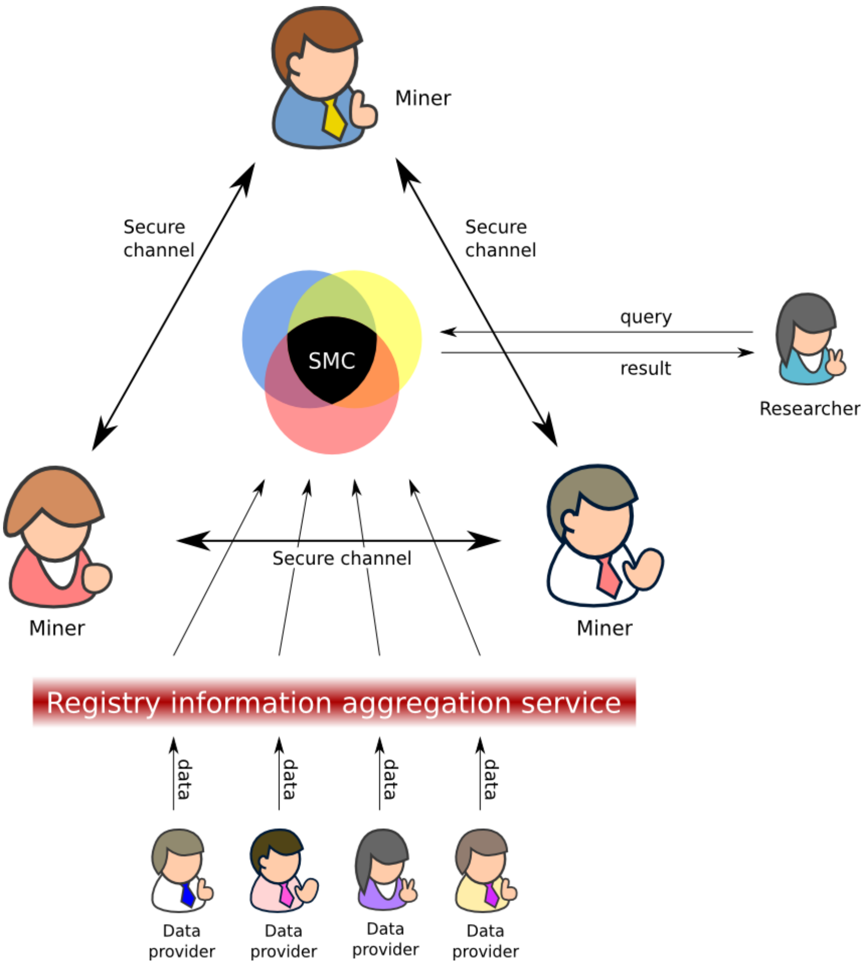

Figure 7. Case D2a: A researcher interested in data from state databases using third-party miners

3.8 Case D2b: Researcher interested in data from state databases using built-in computing possibilities of state infrastructure

Who are data providers?	In this case, there are two layers of data providers. For the SMC, the data are provided from state owned databases. Each of these has also numerous data providers whose interest is in maximum protection of their information and proper data protection laws being reinforced.
Who are data miners?	While in all other aspects, this solution is very similar to the previous one, with the availability of the proper infrastructure, there is no need to give out data to trusted third parties. Instead, each database has the computing capabilities through state infrastructure.
What are the challenges for the trust?	Challenges for the trust are mostly in proper query control
How is the data collated?	Vertically
Who can submit the query?	Interested researcher from outside
How is responsibility for the query distributed?	Legislatively regulated responsibility for checking the adequacy of the query
Who get to know the results?	The interested researcher and checking authorities
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel
What is the motivation for computing and keeping the trust?	SMC enables to put state databases into maximum use for research or policy making purposes, while maintaining security and privacy of each database and each data provider for those databases and avoiding merge of actual data. Data security officers need to approve of the security and legality of the solution.
Illustration	Figure 8. The registry information aggregation service is SMC capable on its own and the circle represents all the aggregated state-databases where the joint query is addressed.

State owns numerous databases, but state is legally not allowed to merge the databases for any joint computation. We look at an interested researcher or policy maker, who is with some authorization, allowed to make a query that gives a general answer across multiple databases, but computations that enable case-based identification are not allowed. This model assumes that computation capabilities can be built in to common registry aggregation service (e.g. X-road for Estonia).

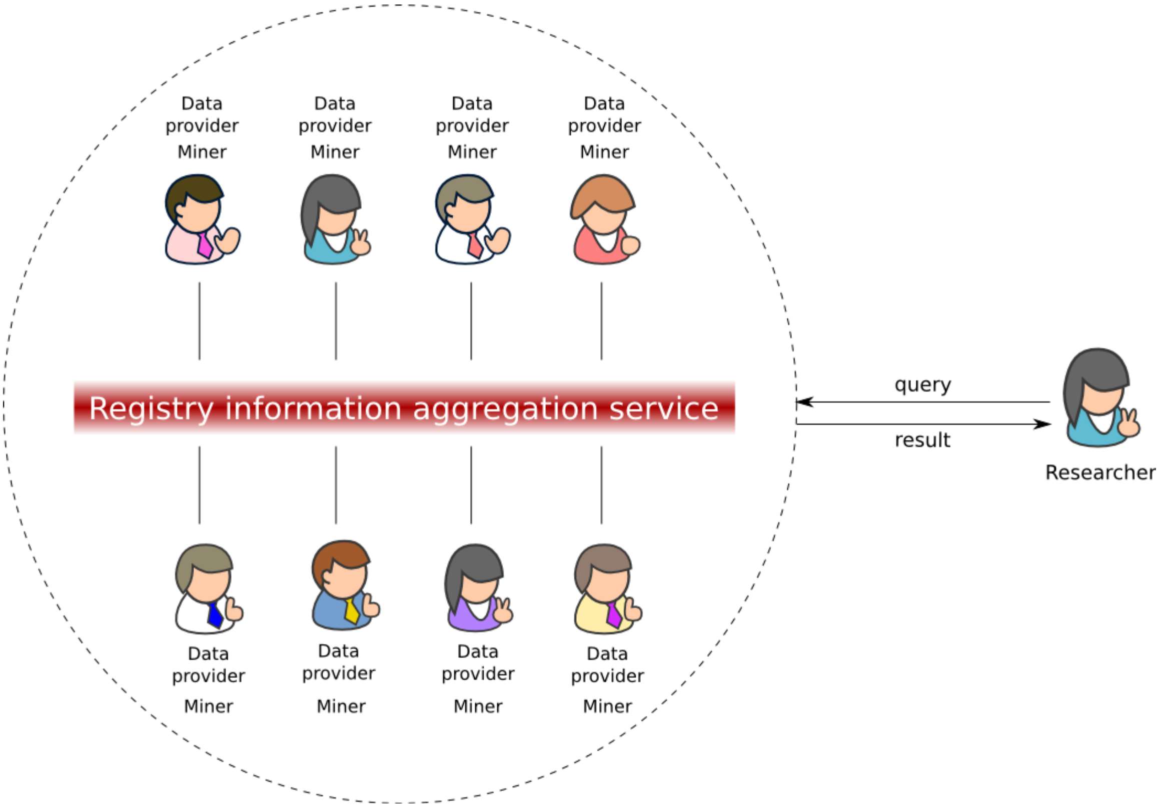

Figure 8. Case D2b: A researcher interested in data from state databases using built-in computing possibilities

3.9 *Case D3: Statistical data collection organizations working together for better results*

Who are data providers?	Numerous – everyone participating in a survey. The survey data can be collected through web-based self-completed surveys, or computer assisted personal interviews or computer assisted telephone interviews.
Who are data miners?	Three competing parties
What are the challenges for the trust?	Two parties working together. In addition, proper query control is needed so that only sections of information could be analysed.
How is the data collated?	Horizontally. Different parties can request variables to be included in the data collection and they share a common set of variables.
Who can submit the query?	One party with authorization from others or with automatic authorization rules that regulate data query rights.
How is responsibility for the query distributed?	Authorization of other parties is needed to make sure that the query is made only so that the section of information is received that has been "paid for".
Who get to know the results?	One party
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel. Data collection channel
What is the motivation for computing and keeping the trust?	Reduction of costs, increasing statistical relevance
Illustration	Figure 9

Statistical data are viable the more respondents are available, however, each respondent is fairly expensive, thus there is an optimal error margin which is accepted as reasonable. If different statistical data gathering market research companies join forces and share the interests of different data collection tasks (e.g. omnibus surveys), it is possible to reduce cost of data collection by sharing it between different organizations. SMC enables to keep data from the competing marketing company, while joining resources for data collection.

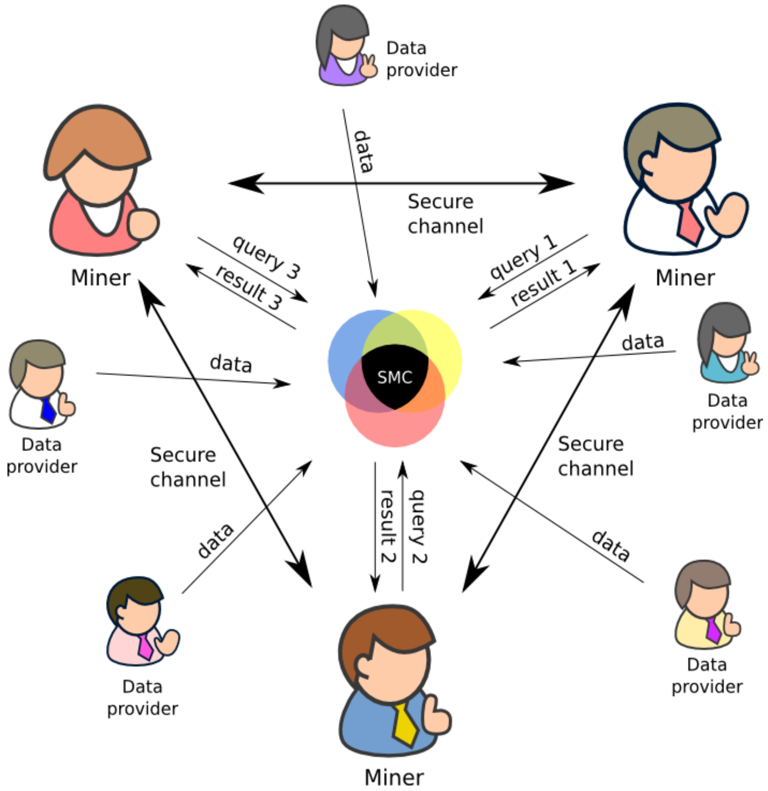

Figure 9. Case D3: Statistical data collection organizations working together for better results

3.10 Case D4: Statistical office collecting data with improved security

Who are data providers?	Numerous – everyone participating in a survey or everyone who has legal obligations to provide data. The survey data are provided through web-based self-completed surveys.
Who are data miners?	One party with two outsourced data hosting SMC capable organizations for additional security assurance for the data provider.
What are the challenges for the trust?	Two parties working together against the data owner.
How is the data collated?	Horizontally
Who can submit the query?	Data owner. Other parties are not allowed to submit queries.
How is responsibility for the query distributed?	Data owner's possibilities for query come from legislations.
Who get to know the results?	Data owner
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel. Data collection channel.
What is the motivation for computing and keeping the trust?	Increasing security and trustworthiness among those who need to provide the data.
Illustration	Figure 10

Many entities need to provide statistical data to state statistics office. The legal rules declare the way and principles of data provision, SMC provides additional confirmation that the anonymity is protected and that the data are kept secure.

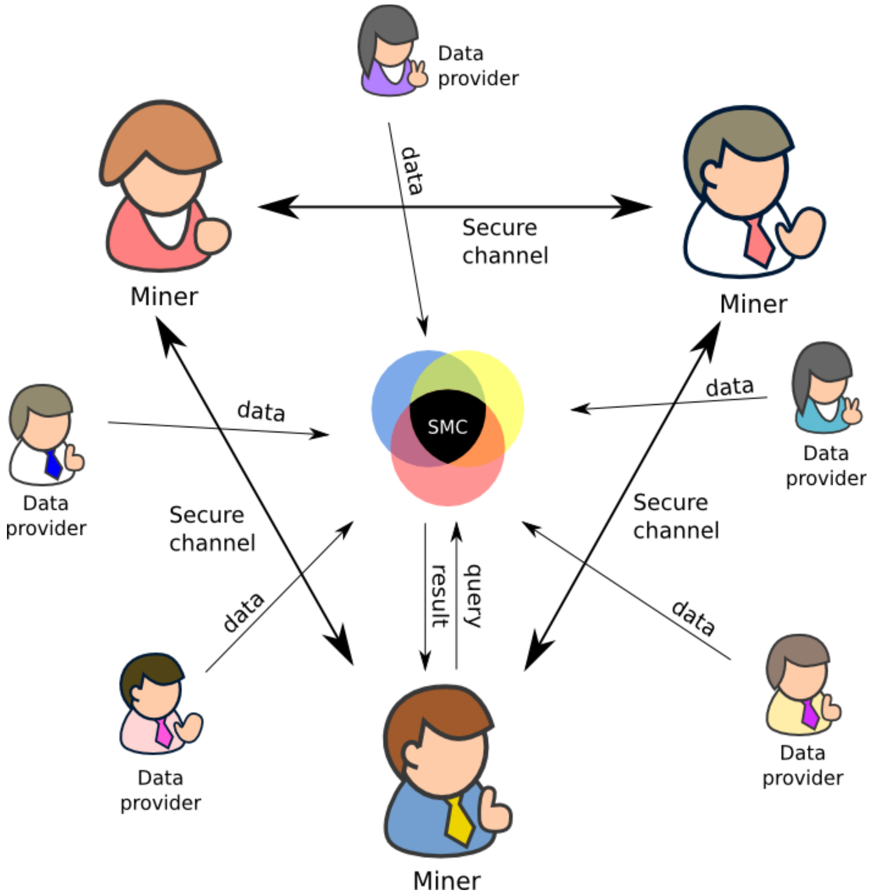

Figure 10. Case D4: Statistics office collecting data with improved security

3.11 Case E1: State database and interested non-state parties

Who are data providers?	At least three interested parties – one state database and two or more non-state entities with similar data.
Who are data miners, the ones who need to compute for the solution?	State data provider and two non-state providers with SMC capabilities. Case E2 is a version of this, where computing capabilities are outsourced.
What are the challenges for the trust?	Two parties working together
How is the data collated?	Vertically
Who can submit the query?	All parties in agreement
How is responsibility for the query distributed?	All parties in agreement, with additional claims from the state
Who get to know the results?	All involved parties
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel; Data collection channel
What is the motivation for computing and keeping the trust?	To provide information about possible competitive edge
Illustration	Figure 11. The illustration is the same as for case C (Figure 5). The difference of the two cases lies mainly in the way data is collated. The circle depicts numerous data providers belonging to the same organization

Consider an example where several universities want to know the "value" of their education by analysing the net revenues of their graduates. Then joining tax office databases with graduate databases would provide relevant information.

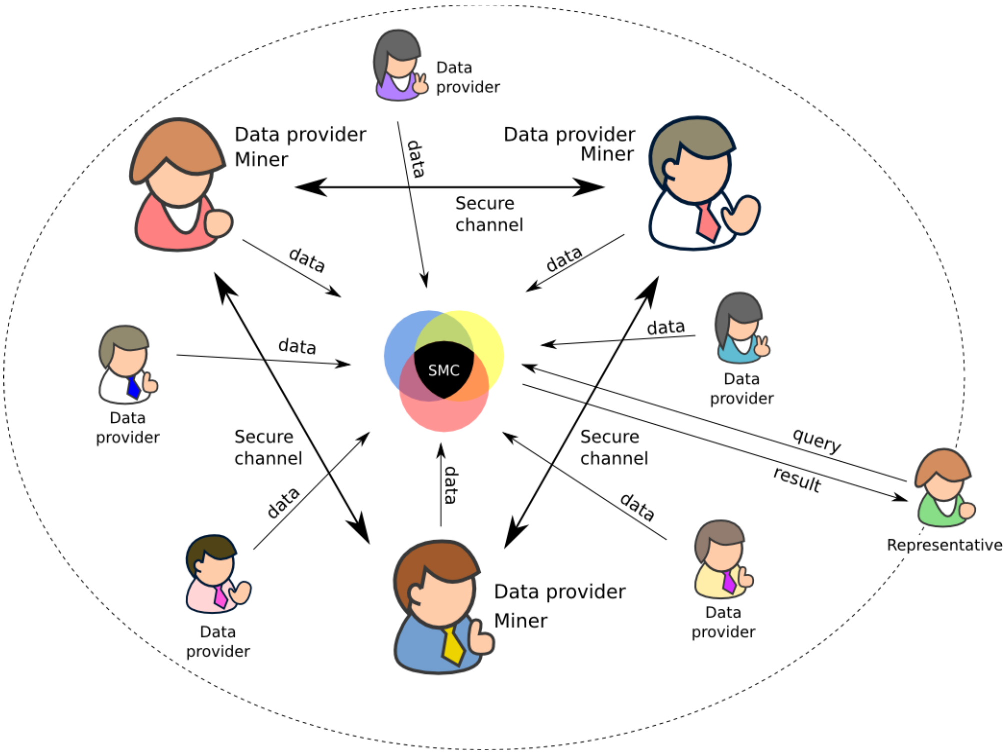

Figure 11. Case E1: State database and interested non-state parties

3.12 Case E2: State database and interested non-state parties (with the computation outsourced)

Who are data providers?	At least three interested parties – one state database and two or more non-state entities with similar data
Who are data miners?	Computing capabilities are outsourced
What are the challenges for the trust?	Two parties working together
How is the data collated?	Vertically
Who can submit the query?	All parties in agreement
How is responsibility for the query distributed?	All parties in agreement, with additional claims from the state
Who get to know the results?	All involved parties
What are the possibilities for security attack?	Communication channels; SMC model; Simultaneous attack on two parties; Simultaneous attack on one party and the opposite communication channel; Data collection channel
What is the motivation for computing and keeping the trust?	To provide information about possible competitive edge. SMC computation is outsourced, but the trust is that the miners are not interested in disclosing other data, as it would break their trustworthiness on the market.
Illustration	Figure 12. The circle represents the fact that a joint representative from all data providers is needed for this case

Here the use-case is similar to case E1, but the computation is outsourced to third parties.

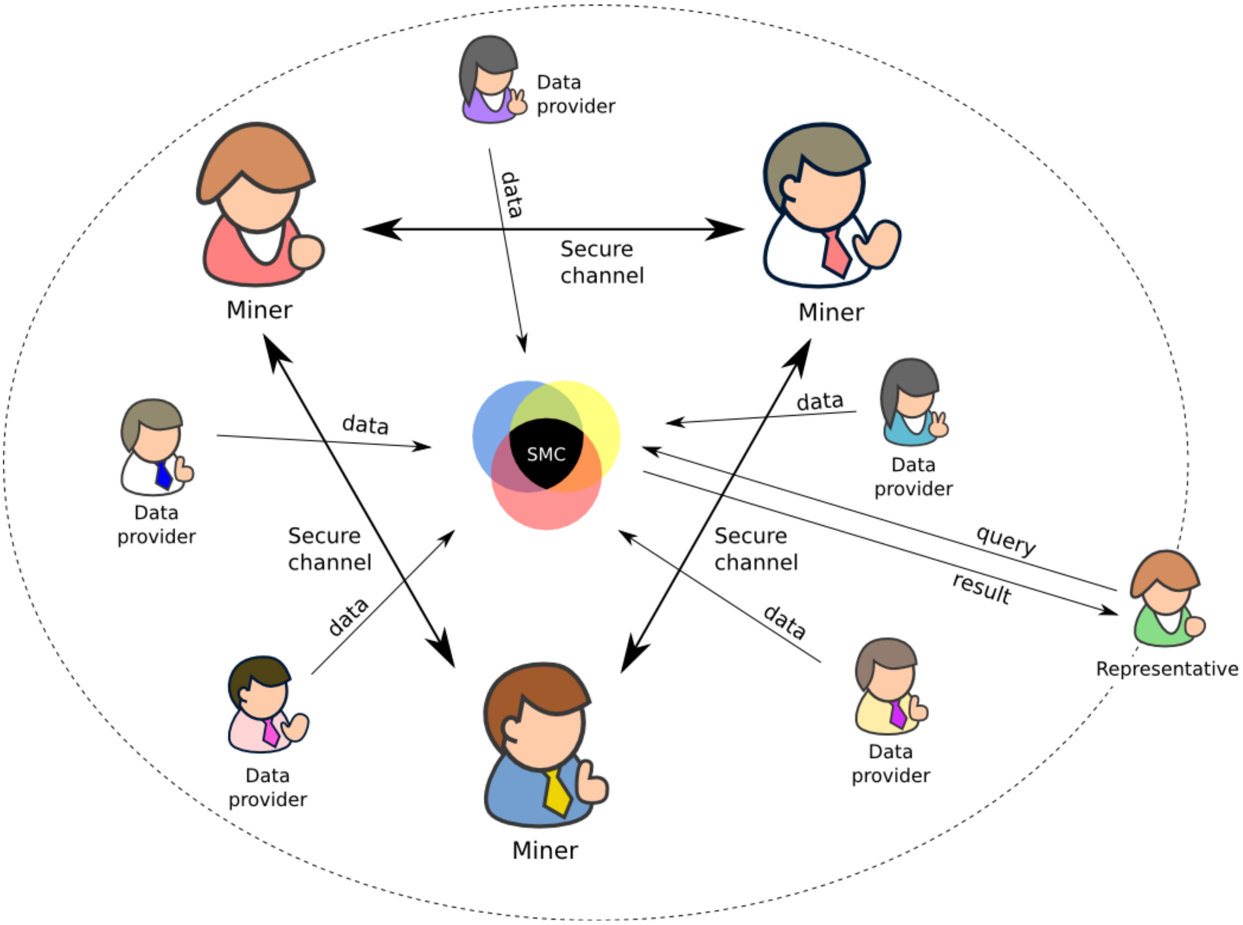

Figure 12. Case E2: State database and interested non-state parties (with the computation outsourced)